ASNR AMERICAN SOCIETY OF NEUROREHABILITATION

Neurorehabilitation and Neural Repair
1–11

Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/15459683261469063
journals.sagepub.com/home/nnr
Sage

Applying the International Classification of Functioning, Disability and Health to Classify Upper Limb Outcome Measures in Stroke Rehabilitation: A Construct-Driven Perspective

Canan Yuksel-Haji, PhD[1,2], Nicola Hancock, PhD[1], Allan Clark, PhD[3], and Valerie M. Pomeroy, PhD[1,4]

Abstract

Introduction: Accurate evaluation of upper limb recovery after stroke is essential for guiding rehabilitation targeting specific impairments and activity limitations. To achieve this, outcome measures must clearly reflect what they assess, whether neuromuscular impairments or activity limitations. However, classification of measures is not always clear-cut, which can lead to misinterpretation of recovery progress and reduce rehabilitation planning effectiveness.

Problem: Current classification practices often prioritize item content over the scoring criteria that define what is being quantified. This becomes particularly problematic when task-based items are scored using impairment-oriented criteria (e.g., range of motion), blurring whether change reflects impairment recovery or activity-level improvement. This ambiguity can mislead clinical decision-making and the interpretation of research findings.

Solution: We propose a construct-driven application of the International Classification of Functioning, Disability and Health (ICF) in which component-level attribution is determined by the construct operationalized through scoring rather than inferred from item content. As scoring structures differ, three strategies are suggested based on item-level scoring attribution where feasible: (1) proportional classification for measures in which one ICF component predominates (with a pragmatic, provisional ≥60% dominance guideline for near-boundary cases), otherwise (2) dual-component reporting for mixed measures; and (3) non-proportional interpretation for measures exhibiting construct shifting across score levels or using dominant-construct scoring with embedded grading criteria.

Conclusion: Grounding classification in the construct operationalized by scoring criteria provides a more defensible basis for outcome selection and interpretation. This supports clearer interpretation of intervention effects, more targeted rehabilitation, and stronger comparability across trials and evidence synthesis.

Keywords

stroke rehabilitation, outcome measures, motor impairment, neuromuscular impairment, activity limitation, upper limb, clinical assessment

Introduction

There is international consensus on the need for standardized, conceptually grounded outcome measures of upper limb (UL) neuromuscular impairments and activity limitations in stroke rehabilitation.[1-3] But do outcome measures in widespread use accurately capture the constructs they intend to assess? Specifically, do measures classified as assessing neuromuscular impairment or activity limitation do so with precision? These questions are not trivial. Accurate measurement is fundamental to understanding, guiding, and improving UL recovery for people after stroke.

Improving UL recovery, one of the top global research priorities in stroke rehabilitation,[4,5] is especially important

[1]School of Health Sciences, University of East Anglia, Norwich, UK
[2]Faculty of Health Sciences, Sivas Cumhuriyet University, Sivas, Türkiye
[3]Norwich Medical School, University of East Anglia, Norwich, UK
[4]Department of Clinical Neurosciences, University of Cambridge, UK; NIHR HealthTech Research Centre in Brain Injury, UK; Professor Emeritus, University of East Anglia, Norwich, UK

Corresponding Author:
Canan Yuksel-Haji, Faculty of Health Sciences, Sivas Cumhuriyet University, Sivas, Türkiye.
Email: cananyuksel58@gmail.com

given the high prevalence of UL neuromuscular impairments after stroke, affecting approximately 35% to 57% of survivors.[6,7] In parallel, activity limitations remain highly prevalent after stroke, underscoring the functional impact of neuromuscular impairments on everyday task performance.[8,9] These limitations have profound consequences for independence, participation, and return to work.[10] Therefore, identifying interventions that effectively target UL neuromuscular impairments is seen as essential.[11] Critically, this first requires outcome measures that can clearly differentiate between impairments and activity limitations, allowing clinicians and researchers to determine whether an intervention influences neuromuscular impairments, activity limitations, or both independently.

Accordingly, the present paper focuses on neuromuscular impairments and activity limitations, where construct ambiguity is most likely to arise in UL measures, particularly when task-based items are scored using impairment-oriented criteria (e.g., voluntary movement control and coordination). Participation restrictions are not considered here because participation outcomes are strongly shaped by contextual factors (environmental and personal), and addressing them would require additional examples and interpretive decisions beyond the scope of this paper. For clarity, the term neuromuscular impairment is used in this paper to denote the same underlying construct as motor impairment, while retaining emphasis on the neuromuscular processes underlying movement.[12]

Problem

Many UL outcome measures are routinely classified as assessing either neuromuscular impairments or activity limitations. In practice, however, some widely used tools include elements of both constructs despite being labeled as one or the other.[13,14] This overlap makes it difficult to determine what aspect of UL recovery a given measure really captures and weakens confidence in how change is interpreted. The consequences are practical: classification ambiguity can distort conclusions about recovery processes, reduce comparability across studies, and complicate decisions about which outcomes best match intervention targets. Without a more stringent and consistent approach to classifying UL measures, both clinical practice and research risk drawing conclusions from scores that do not map clearly onto the constructs they are assumed to represent.

Utilizing the ICF Framework for Precise Classification of Measures

The International Classification of Functioning, Disability, and Health (ICF) framework offers a conceptually grounded and widely recognized structure for classifying outcome measures by linking them to ICF categories, which are organized within the components of Body Functions and Structures, and Activities and Participation, along with their corresponding negative aspects (impairments, activity limitations, and participation restrictions).[15] In theory, it can support more accurate classification and clearer distinctions between what outcome measures are intended to assess. However, its effective application in practice is often inconsistent.

Misclassifications may arise from misinterpretation of distinctions between ICF components, inconsistent application of linking rules, and difficulties in identifying core concepts within assessment tools.[16] These challenges limit the ICF's potential to resolve existing ambiguities and highlight the need for more stringent and systematic classification approaches.

To understand how the ICF can be applied more effectively in stroke rehabilitation, it is important to briefly revisit how impairments and activity limitations are defined within the framework and how these definitions inform the classification of outcome measures. According to the ICF, impairments refer to problems in body functions, such as reduced muscle strength, increased muscle tone, or impaired voluntary movement control and coordination, and body structures, including joint contractures or a subluxed shoulder.[17] Activity limitations refer to difficulties in executing purposeful tasks, such as lifting or carrying an object and feeding oneself.[17] In the context of UL neuromuscular recovery post-stroke, impairments in body functions related to the neuromuscular system are referred to as neuromuscular impairments.[17,18] Neuromuscular impairment measures assess the physiological and biomechanical functions that may underlie activity limitations, whereas activity measures (i.e., measures primarily capturing activity limitations) evaluate how well individuals perform daily tasks after stroke, considering time, effort, and level of assistance required (Table 1).[15]

Challenges with the Classification of Outcome Measures

The ICF framework has conceptual clarity, but practical applications have typically focused on the content of a measure's items.[13,19-21] This approach, however, can be misleading when the construct operationalized through scoring criteria differs from that suggested by item content.

ICF-based classification is commonly guided by the linking rules proposed by Cieza and colleagues,[21,22] which recommend identifying "meaningful concepts" within items and linking them to the most precise ICF categories. Importantly, item-level linking guides interpretation but does not, on its own, specify what is quantified; this depends on how performance is operationalized through scoring. Consequently, in task-based UL measures, the item content (e.g., "activity") does not always align with the construct quantified by the scoring criteria.

Table 1. Specific Domains of Interest Assessed by Neuromuscular Impairment and Activity Measures.

Neuromuscular impairment measures	Activity measures
• Underlying neuromuscular characteristics that constrain movement production (e.g., tone, strength, voluntary control, coordination) • Quality of movement operationalized through impairment-level scoring criteria (e.g., synergy, smoothness, voluntary movement control), independent of task accomplishment • Deviation of movement patterns from typical/normative movement characteristics, even when a task can be completed	• Which activities/tasks can or cannot be accomplished (task success/independence) • Time required to complete a functional activity/task (performance speed) • Level of assistance required to complete an activity/task (human help/device support) • Effort required for task completion (e.g., number of attempts, time cost, perceived difficulty)

Conceptual grouping informed by the International Classification of Functioning, Disability and Health and the Stroke Recovery and Rehabilitation Roundtable recommendations.[1,15] "Activity measures" refers to measures primarily capturing activity limitations.

The "Motor Evaluation Scale for Upper Extremity in Stroke Patients (MESUPES)," for instance, has a stated aim of assessing UL movement quality.[23] Its items include both direct evaluations of body functions (e.g., shoulder abduction, finger extension) and task-based actions involving object interactions (e.g., gripping a plastic bottle or dice).[23-25] If classification is guided primarily by item wording, the task-based content may suggest an activity-level label, even though the scoring criteria point to an impairment-level construct. In MESUPES, this is evident in the scoring structure: the arm subscale grades tone adaptation, muscle contraction and joint movement, while the hand subscales assess hand range of motion (ROM) and hand orientation during task performance.[26] Accordingly, although some items are activity-like in content, MESUPES predominantly quantifies body functions (neuromuscular impairment), supporting its classification as a neuromuscular impairment measure and illustrating how reliance on item content alone, rather than what the scoring rewards, can misclassify a measure's primary construct (Table 2).

Another challenge arises when scoring criteria incorporate both impairment-level and activity-level constructs within the ICF framework.[19] For instance, although the Wolf Motor Function Test (WMFT) is widely used and commonly classified as an activity measure,[27-29] aspects of its Functional Ability Scale have also been discussed in relation to neuromuscular impairment.[30] The scale incorporates both impairment- and activity limitation-related elements, and its scoring criteria shift the construct being quantified across the 6 score levels (0-5).

Scores of 0 (no attempt), 1 (non-functional attempt), and 2 (task attempted with assistance)[31,32] primarily reflect activity limitation because they capture whether and how independently a task can be accomplished rather than specifying movement quality. A score of 3, awarded for movements performed with synergy, slowness, or excessive effort, reflects a mixed construct because it includes both activity performance (e.g., task completion) and neuromuscular impairment-related movement features (e.g., synergy). In contrast, scores of 4 and 5, assigned to movements that increasingly approximate pre-stroke patterns, predominantly reflect neuromuscular impairment, as they are determined by movement quality, coordination, and normality rather than task success. Thus, while the WMFT is commonly classified as an activity measure, its Functional Ability Scale operationalizes both neuromuscular impairment and activity limitation constructs, corresponding to the ICF Body Functions and Structures, and Activities and Participation components, respectively. This illustrates how reliance on item content alone can obscure the construct quantified by scoring (Table 2).

Table 2 illustrates how ICF linking based on item content versus the construct operationalized by scoring criteria may yield different category-level links and, consequently, different component-level attribution.

Taken together, the examples in Table 2 underscore the complexity of mapping outcome measures onto the ICF framework, especially where the construct operationalized by scoring diverges from item content or where measures encompass multiple constructs corresponding to different ICF components. Under these conditions, classifications derived primarily from item wording risk being inaccurate and may fail to reflect the construct evaluated by the measure. A more systematic and standardized classification process is therefore required to improve the interpretability of research outcomes and enhance clinical applicability.

Solution

Accurate classification of stroke outcome measures is essential, as confusion about what a measure truly assesses (e.g., neuromuscular impairment versus activity limitation) can lead to misinterpretation at multiple levels.[19,20,23,33,34] If an outcome is labeled as an activity measure but its scoring primarily captures neuromuscular impairment (or vice versa), treatment effects may be misattributed, suggesting improvements in task execution when changes primarily reflect impairment-level recovery, or masking genuine gains in activity-level functioning when impairment-focused scoring dominates. Such misinterpretation can directly affect clinical

Table 2. Content-Based Versus Scoring-Criteria-Based ICF Linking and Resulting Component-Level Attribution of Selected Measures.

Measure	Item content	ICF link(s) suggested by item content	Scoring criteria	ICF link(s) suggested by the construct operationalized by scoring
MESUPES - Arm	**Items:** 1. Hand to stomach 2. Hand back to the starting position 3. Abduction 0°-90° (arm slides on the table) 4. Arm back to the starting position 5. Hand from knee onto the table 6. Hand (palm) to mouth (elbow on the table) 8. Hand on top of the head (shoulder in abduction) 7. Reach with correct orientation of fingers and wrist (as if to grasp) for a plastic bottle (grasp not required)	**Items 1-6 and 8:** **d445** Hand and arm use **Item 7:** **d445** Hand and arm use **d4452** Reaching **d4401** Grasping (secondary—"as if to grasp", but grasping not required)	**0** = No adequate adaptation of tone to the movement (hyper-or hypotonia) **1** = Adequate (normal) tone adaptation to at least part of the movement **2** = Participation through normal muscle contraction in at least part of the movement (assisted as needed to perform the movement normally) **3** = Performs part of the whole movement normally (without help) **4** = Completes the whole movement normally but performs it slowly or with great effort **5** = Completes the whole movement normally at normal speed	**Scores 0-1:** **b7350** Tone of isolated muscles and muscle groups **Score 2:** **b730** Muscle power functions (normal muscle contraction) **b760** Control of voluntary movement functions (volitional contribution) **Scores 3-5:** **b760** Control of voluntary movement functions **b7602** Coordination of voluntary movements **b710** Mobility of joint functions **b730** Muscle power functions
MESUPES - Hand / A. ROM	**Items:** 1. Pinch grip (thumb-index opposition) 2. Wrist extension 3. Opposition thumb-little finger 4. Selective extension of the third finger 5. Index-middle finger abduction 6. Selective extension of the fifth finger	**Item 1:** **b710** Mobility of joint functions (primary) **d4401** Grasping (secondary; implied by "pinch grip") **Item 2, 3, 4, 5 and 6:** **b710** Mobility of joint functions	**0** = No movement **1** = Movement amplitude $<$ 2 cm **2** = Movement amplitude $\geq$ 2 cm ("Movement amplitude" is the construct measured in the ROM test.)	**Scores 0-2:** **b710** Mobility of joint functions (movement amplitude/ROM)
MESUPES - Hand / B. Orientation	**Items:** 1. Grip the bottle (thumb-index pinch) and lift 2 cm 2. Grip dice (thumb-index pinch) and rotate once (vertical axis) 3. Index fingertip on dice; rotate once (vertical axis) using fingers 1 and 3	**Item 1:** **d4401** Grasping **d4402** Manipulating **d4300** Lifting **Item 2:** **d4401** Grasping **d4402** Manipulating **Item 3:** **d4402** Manipulating	**0** = No movement or abnormal finger/wrist orientation toward the object **1** = Normal orientation of either fingers or wrist toward the object **2** = Whole movement correct (normal orientation throughout)	**Scores 0-2:** **b760** Control of voluntary movement functions **b7602** Coordination of voluntary movements

(continued)

Table 2. (continued)

Measure	Item content	ICF link(s) suggested by item content	Scoring criteria	ICF link(s) suggested by the construct operationalized by scoring
WMFT - Time + Functional Ability Scale (FAS)	**Timed - and FAS-scored items:** 1. Forearm to table (side) 2. Forearm to box (side) 3. Extend elbow (side) 4. Extend elbow (weight) 5. Hand to table (front) 6. Hand to box (front) 8. Reach and retrieve 9. Lift can 10. Lift pencil 11. Lift paper clip 12. Stack checkers 13. Flip cards 15. Turn key in lock 16. Fold towel 17. Lift basket	**Items 1–6:** **d445** Hand and arm use **Item 8:** **d445** Hand and arm use **d4452** Reaching **Items 9–11:** **d4300** Lifting **d4401** Grasping **Items 12–13:** **d4402** Manipulating **Item 15:** **d4402** Manipulating **d4453** Turning/twisting hands/ arms **Item 16:** **d4402** Manipulating **Item 17:** **d4300** Lifting **d4401** Grasping **d4301** Carrying in the hands	**Time (s):** Time required to complete each task **FAS (0–5):** **0** = No attempt **1** = Attempt but no functional participation **2** = Performs with assistance or requires >2 attempts or is very slow **3** = Performs; influenced by synergy or performed slowly or with effort **4** = Performs; movement close to normal but slightly slower and/or lacking precision, fine coordination or fluidity **5** = Performs; movement appears normal	**Time score (seconds):** **Activities and Participation**—Linked at the component level only; scoring primarily reflects speed of task execution. **FAS 0-2:** **Activities and Participation**—Linked at the component level only; scoring primarily reflects task accomplishment and independence level. **FAS 3:** **Mixed construct** **Activities and Participation**—Task completion. **Body Functions and Structures:** **b760** Control of voluntary movement functions **b7602** Coordination of voluntary movements **FAS 4–5:** **Body Functions and Structures (movement quality/normality):** **b760** Control of voluntary movement functions **b7602** Coordination of voluntary movement
WMFT - Quantitative Items:	**Items:** 7. Weight to box (lbs) 14. Grip strength (kg)	**Item 7:** **d4300** Lifting **d445** Hand and arm use **Item 14:** **b730** Muscle power functions	**Weight to box (lbs):** Amount of weight lifted during task performance, recorded in pounds **Grip strength (kg):** Force generated during handgrip assessment	**Weight to box (lbs):** **Body Functions and Structures (Linked at the component level only)**—Quantitative physical capacity output **Grip strength (kg):** **b730** Muscle power functions

Concepts were linked to the most precise ICF categories based on item content and, where applicable, scoring criteria. When categories could not be specified, linking was performed at the component level. Component-level attribution can be inferred from the component to which the linked categories belong. d-codes denote the ICF Activities and Participation component; b-codes denote the Body Functions part of the ICF Body Functions and Structures component. Item content and scoring criteria were summarized and adapted from the MESUPES publication (Van De Winckel, Feys and Van Der Knaap, 2006) and the WMFT manual (Taub, Morris and Crago, 2011); the ICF links and construct-level attributions were developed by the authors.
Abbreviations: WMFT, Wolf Motor Function Test; MESUPES, Motor Evaluation Scale for Upper Extremity in Stroke Patients; ROM, range of motion; FAS, Functional Ability Scale.

decision-making by obscuring which therapeutic target has changed, making it harder to determine whether an intervention is addressing neuromuscular impairment, activity limitation, or both.[19,20,35] Ultimately, this risks misalignment between assessment results and outcomes that matter most to people living with stroke, including meaningful improvements in independence and day-to-day task performance.[19,20]

At a broader level, construct mismatch contributes to heterogeneity across clinical trials and can undermine the validity of evidence synthesis in systematic reviews and meta-analyses.[1,34] Because these secondary outputs are frequently treated as high-level evidence informing clinical guidelines and rehabilitation policy, misclassification risks propagating construct error into recommendations and practice.[35] Improving classification consistency, therefore, supports clearer interpretation of treatment effects, more appropriate outcome selection in trials, and more trustworthy evidence synthesis in stroke rehabilitation.[19,20,36]

To address these challenges in outcome-measure classification, the focus should shift from item content alone to the construct operationalized through scoring criteria. For instance, a task such as reaching for a cup above shoulder level may appear to assess activity limitation given the task performed; however, when scoring is anchored in shoulder ROM, it more accurately reflects neuromuscular impairment. By examining what the scoring is designed to quantify, outcome measures can be more appropriately classified within the ICF framework. The key question, therefore, is whether scoring mainly captures neuromuscular impairments (e.g., strength, ROM, coordination) or activity limitations (e.g., time required, effort, or need for assistance). However, in practice, some measures also incorporate scoring elements that reflect both constructs to varying degrees.

For measures in which scoring reflects both neuromuscular impairment and activity limitation constructs, classification can be more transparently justified by considering the relative contribution of these constructs within the scoring criteria. As a pragmatic guideline, measures may be classified as impairment-focused when approximately 60% or more of the scoring reflects neuromuscular impairments, and as activity-focused when approximately 60% or more reflects activity limitations. This threshold is intended to reduce uncertainty in near-balanced cases (e.g., 51% vs 49%) and to provide an explicit rationale for categorization. Importantly, the cut-off is based on expert judgment and should be treated as a practical starting point to be refined through future empirical validation.

This proportional approach is most appropriate for item-level measures, in which individual items serve as separable indicators of distinct constructs (e.g., impairments versus activity limitations). In such measures, construct representation is distributed at the item level, allowing the relative proportion of impairment-related and activity-related items to be meaningfully quantified and supporting proportional (percentage-based) classification.

This approach is illustrated using the Fugl-Meyer Assessment (FMA) UL Motor Scale as an item-level measure (Table 3). The FMA-UL Motor Scale comprises 33 items across 4 sections (upper extremity, wrist, hand, and coordination/speed), each scored using explicit criteria that operationalize the construct being quantified.[37,38] At the item level, 32/33 items (≈97%) directly operationalize neuromuscular impairment constructs (e.g., reflex activity, selective movement, joint mobility, force generation, coordination), corresponding to the ICF Body Functions and Structures component. The remaining item (≈3%) uses a time-based metric (time difference relative to the unaffected side) during a standardized finger-to-nose coordination probe. Although time can be framed as an activity-performance attribute, here it functions as a performance descriptor within a standardized motor probe rather than an index of functional task accomplishment, independence, or participation, and is thus more appropriately interpreted as reflecting neuromuscular impairment rather than activity limitation. Accordingly, the FMA-UL Motor Scale remains impairment-focused at the scale level, representing a clear dominance case for the proportional approach; thus, the proposed ≥60% guideline is easily met in this case (Table 3).

A second item-level example is the Manual Function Test (MFT; Table 4), which represents a near-boundary case for the proportional approach. The MFT comprises 32 dichotomously scored items divided evenly between constructs[39]: 16 Arm Motions items primarily operationalize graded voluntary movement capacity (neuromuscular impairment; ICF Body Functions and Structures component), whereas 16 Manipulative Activities items operationalize goal-directed task accomplishment (activity limitation; ICF Activities and Participation component). Two manipulation items, carry a cube (CC) and peg-board (PP), also incorporate time-based efficiency constraints. As scoring is evenly distributed (~50% impairment; ~50% activity limitation), the MFT does not meet the proposed 60% threshold for classification as predominantly impairment- or activity-focused. It is therefore best described as an item-level mixed measure with explicit dual-component coverage (i.e., dual impairment and activity limitation constructs).

However, proportional classification is not appropriate for all outcome measures. Some exhibit score-level construct shifting, in which the construct reflected by the score changes across the rating scale (i.e., different score bands prioritize different constructs), rather than being separated across items. In these cases, different parts of the scoring range may reflect distinct emphases (e.g., task completion/assistance at one end versus impairment features at the other), and some score levels may incorporate a mixed construct that captures both. This can make item- or percentage-based classification misleading. These measures are therefore not "mixed" at the item level; instead, construct attribution varies by score level. A score-level interpretive approach may better represent this

Table 3. Scoring-Criteria-Based ICF Component Attribution for the FMA-UL Motor Scale (Single-Component Dominant Measure).

Section	Item set (n)	Scoring criteria	Primary construct evaluated by scoring	ICF component reflected by scoring
A. Upper extremity	Reflex activity (I) (n = 2) Normal reflex activity (V) (n = 1; conditional on max Part IV)	**I: 0** = None; **2** = Can be elicited **V: 0** = Hyperactive; **1** = Lively; **2** = Normal	Reflex responsiveness: elicitation/normality	Body Functions and Structures
	Volitional movement (II–IV): within, mixed, minimal synergy (n = 15)	**0** = None **1** = Partial **2** = Full	Voluntary movement	Body Functions and Structures
B. Wrist	Stability at 15° DF (elbow 90° and 0°) + repeated DF/VF (elbow 90° and 0°) + circumduction (n = 5)	**Stability at 15° DF:** **0** = <15°; **1** = 15° no resistance; **2** = holds vs resistance. **Repeated DF/ VF + Circumduction:** **0** = none; **1** = partial/jerky; **2** = full smooth.	Wrist voluntary movement: ROM, positional hold, and smoothness (where specified)	Body Functions and Structures
C. Hand	Mass flexion/Mass extension (n = 2)	**0** = None **1** = Partial **2** = Full	Gross voluntary hand ROM	Body Functions and Structures
	Grasp patterns (hook, thumb adduction, pincer, cylinder, spherical; object + resistance/ tug) (n = 5)	**0** = Cannot perform **1** = Holds but weak/not against tug or resistance **2** = Holds firmly/against tug or resistance	Specific grasp pattern formation and hold against tug/resistance (Objects function as probes)	Body Functions and Structures
D. Coordination/ speed	Tremor (n = 1)	**0** = Marked **1** = Slight **2** = None	Involuntary movement stability (Tremor severity)	Body Functions and Structures
	Dysmetria (n = 1)	**0** = Pronounced/unsystematic **1** = Slight/systematic **2** = No dysmetria	Coordination accuracy	Body Functions and Structures
	Time (finger-to-nose ×5; as fast as possible; scored vs unaffected side) (n = 1)	**0** = ≥6 s slower **1** = 2-5 s slower **2** = <2 s difference	Relative execution speed in a coordination probe (Time is a performance descriptor, not a measure of functional task completion).	Body Functions and Structures

Scoring-criteria-based attribution in an item-level measure: FMA-UL Motor Scale. Component-level links are derived from the construct operationalized by the scoring criteria. Across items, scoring maps to the ICF Body Functions component (b-domain neuromusculoskeletal functions), reflecting neuromuscular capacity (impairment-level construct). Item sets and scoring criteria were adapted from the FMA publication (Fugl-Meyer et al., 1975)[37] and the standardized FMA-UL manual (Hervé-Colas et al., 2026)[38]; all construct interpretations and ICF component attributions are original to the authors.
Abbreviations: FMA, Fugl-Meyer Assessment; UL, upper limb; DF, dorsiflexion; VF, volar flexion; ROM, range of motion; s, seconds.

structure (e.g., the WMFT Functional Ability Scale; Table 2). Accordingly, total-score change should be interpreted with caution, as improvement at different points on the scale may reflect changes in different underlying constructs.

Additionally, some outcome measures, such as the Action Research Arm Test (ARAT; Table 5), are best classified by a dominant construct rather than by proportional item-splitting. The ARAT is primarily an activity-level measure of UL functional task accomplishment, and its 0-3 ratings represent ordered levels of task completion, with time and qualitative performance features used to distinguish adjacent score categories.[40] These features (e.g., movement quality, speed, and apparent effort) operate as embedded grading criteria within the same activity construct and do not warrant reattributing the score to separate constructs. While the ARAT literature does not formally adopt ICF "qualifier" terminology, the scoring logic is conceptually similar to the ICF category-qualifier distinction in which a category is identified and

Table 4. Scoring-Criteria-Based ICF Component Attribution for the Manual Function Test (MFT) (Dual-Component Measure).

Section	Item type	Scoring criteria	Primary construct evaluated by scoring	ICF component reflected by scoring
Arm motions (FE)	Graded shoulder flexion with the elbow extended	Success [1]/failure [0] at 4 ordered angle anchors: <45°, 45-89°, 90-134°, ≥135°	Active shoulder flexion capacity (ROM, angle anchored)	Body functions and structures
Arm motions (LE)	Graded shoulder abduction with elbow extended	Success [1]/failure [0] at 4 ordered angle anchors: <45°, 45-89°, 90-134°, ≥135°	Active shoulder abduction capacity (ROM, angle anchored)	Body functions and structures
Arm motions (PO)	Posterior head reach (occiput/temple region)	Success [1]/failure [0] at 4 ordered reach/contact anchors: (1) slight movement/effort; (2) hand above xiphisternal plane; (3) part-hand contact (occiput/temple); (4) palm firmly on occiput + MCPs to occipital midline	Composite UL movement capacity (voluntary movement control, anatomical contact-anchored)	Body functions and structures
Arm motions (PD)	Touch the dorsum with the palm	Success [1]/failure [0] at 4 ordered reach/contact anchors: (1) slight shoulder movement/effort; (2) part-hand contact on ipsilateral buttock; (3) fingertips contact dorsum midline; (4) palm firmly on dorsum + MCP joints extend beyond midline	Composite UL movement capacity (voluntary movement control, anatomical contact-anchored)	Body functions and structures
Manipulative activities (GR)	Ball grasp-hold up-release/drop	Success [1]/failure [0] at 3 ordered task anchors: (1) assisted grasp and maintains hold during active/passive forearm lift; (2) releases after holding up (drops); (3) independent grasp on table→holds up→drops	Object-handling task accomplishment (success + assistance/independence)	Activities and participation
Manipulative activities (PI)	Pinch/pick-up task	Success [1]/failure [0] at 3 ordered object anchors: pick up a pencil; pick up a coin; pick up a needle	Fine object manipulation task success (graded by object demands)	Activities and participation
Manipulative activities (CC)	Cube transfer task (grasp, carry >10 cm beyond the board, release), performed as fast as possible (5 s window)	Success [1]/failure [0] at 4 ordered output anchors defined by cubes completed within 5 s: 1-2, 3-4, 5-6, 7-8	Time-window task output (number completed within 5 s)	Activities and participation
Manipulative activities (PP)	Timed peg insertion (transfer and insert pegs in top-to-bottom order), performed as fast as possible (30 s window)	Success [1]/failure [0] at 6 ordered output anchors defined by pegs correctly inserted within 30 s: 1-3, 4-6, 7-9, 10-12, 13-15, ≥16	Time-window task output (number completed within 30 s)	Activities and participation

Scoring-criteria-based attribution in an item-level measure: MFT. Component-level links are derived from the construct operationalized by the scoring criteria. Items map to ICF Body Functions (b-domain) and Activities and Participation (d-domain) components, with approximately equal representation (~50% impairment-oriented; ~50% activity-level performance). Item types and scoring criteria were summarized and adapted from the MFT manual (Nakamura and Moriyama, 2020); the primary construct interpretations and ICF component attributions were developed by the authors.
Abbreviations: MFT, Manual Function Test; FE, forward elevation; LE, lateral elevation; PO, palm to occiput; PD, palm to dorsum; GR, grasp; PI, pinch; CC, carry a cube; PP, peg-board; MCP, metacarpophalangeal.

Table 5. Scoring-Criteria-Based ICF Component Attribution for the Action Research Arm Test (ARAT) (Dominant-Construct Measure with Embedded Grading Criteria).

Feature	Description
Subtests	Grasp, grip, pinch, gross movement
Item/task content	Standardized, goal-directed UL tasks involving object interaction (e.g., grasp-release, transport, placement) and/or limb placement (gross movement items).
ICF link suggested by item content	Predominantly Activity (d-codes): chiefly d440 (Fine hand use: grasping/manipulating) and d445 (Hand and arm use); item-specific secondary d-codes may apply where relevant.
Scoring anchors (applies to all items)	0 = cannot perform any part of the task (no movement/no component completed); 1 = partial completion (task goal not achieved); 2 = completes the task but with great difficulty and/or in an abnormally long time; 3 = completes the task normally (within the specified time criterion where applied).
Role of neuromuscular features in scoring	Time and effort cues and other observable performance features (e.g., apparent coordination/ smoothness) function as grading cues (especially distinguishing 2 vs 3) within the same activity rating, rather than being scored as a separate independent construct.
Scoring-based ICF component attribution	Dominant construct = Activity (task-goal attainment). Scores 0–3 grade the severity of activity limitation (from inability to normal performance), without using ICF qualifier labels.

Across items, scores 0-3 grade activity-level task accomplishment. Embedded time/difficulty/observable performance cues function as within-construct grading criteria (especially distinguishing 2 vs 3), rather than separate, independent constructs. Therefore, ARAT reflects a dominant activity construct. Subtests, item/task content, and scoring anchors were summarized and adapted from the published standardized ARAT approach (Yozbatiran, Der-Yeghiaian and Cramer, 2008); the ICF interpretations and construct attributions were developed by the authors.

then qualified by extent/severity.[17] Accordingly, the ARAT and similar measures should be classified according to their dominant construct, with grading criteria reported explicitly; proportional classification is therefore conceptually inappropriate in this context.

Explicitly distinguishing between these scoring structures can provide a principled basis for ICF-based classification of outcome measures and help avoid misclassification of the underlying construct within ICF components.

Conclusion

Accurate classification of UL outcome measures is critical because it determines how change is interpreted after stroke, whether an intervention has improved neuromuscular impairment, activity limitation, or both. When measures are misclassified, clinicians may misread progress and tailor rehabilitation toward the wrong target; trial results may be misattributed to the wrong mechanism of recovery; and evidence synthesis may pool non-equivalent constructs, increasing heterogeneity and weakening the interpretability of meta-analyses and guideline recommendations. From a lived-experience perspective, construct clarity supports outcome selection aligned with patient priorities, enabling clinicians to match rehabilitation targets to patient goals, deliver more personalized interventions, and optimize the real-world value of therapy.

To reduce construct ambiguity, we propose an ICF application driven by the construct operationalized through scoring criteria, rather than item content alone. The examples demonstrate that different scoring structures require different interpretive strategies. For item-level scales, proportional classification is appropriate when one construct predominates (e.g., FMA-UL Motor Scale); we propose a pragmatic ≥60% dominance guideline to support transparent decisions in near-balanced cases (noting that this threshold is provisional and requires validation). Near-balanced item-level mixed measures should be reported explicitly as dual-component, reflecting both impairment and activity limitation constructs, rather than forced into a single component (e.g., MFT). For score-level construct-shifting measures (e.g., WMFT-FAS), a single ICF component label for the total score is not conceptually defensible because different score levels operationalize different constructs; interpretation should therefore be score-level aware. Finally, for dominant-construct measures with embedded grading criteria (e.g., ARAT), proportional weighting is conceptually inappropriate; classification should follow the dominant construct, and embedded grading criteria should be reported explicitly as modifiers of performance level rather than treated as separate, independent constructs.

This structured framework has limitations. The proposed ≥60% guideline applies only to item-level scales and should be treated as a pragmatic aid rather than a definitive rule; classifications may be sensitive to item counting/weighting and to how scoring criteria are interpreted across raters, settings, or measure versions. More broadly, scoring-criteria-based construct attribution within the ICF framework still requires transparent judgment where criteria blend constructs or where attribution is feasible only at a coarse (component-level) resolution. Future work should evaluate inter-rater agreement and practical feasibility in criterion-to-component linking and refine the ≥60% guideline through empirical testing and structured expert consensus (e.g., Delphi/panel agreement) involving measurement specialists and clinicians.[41] Despite these limitations, grounding classification in the construct operationalized through scoring criteria can provide a transparent and reproducible basis for outcome selection and interpretation. This, in turn, can reduce construct mismatches and support clearer rehabilitation targeting (impairment vs activity limitation), stronger comparability across trials, and more trustworthy evidence synthesis and guideline recommendations.

Abbreviations

UL: Upper Limb
ICF: International Classification of Functioning, Disability and Health
WMFT: Wolf Motor Function Test
MESUPES: Motor Evaluation Scale for Upper Extremity in Stroke Patients
ROM: Range of Motion
FMA: Fugl-Meyer Assessment
MFT: Manual Function Test
ARAT: Action Research Arm Test

ORCID iDs

Canan Yuksel-Haji https://orcid.org/0000-0002-2645-3855
Nicola Hancock https://orcid.org/0000-0003-4850-3152
Allan Clark https://orcid.org/0000-0003-2965-8941
Valerie M. Pomeroy https://orcid.org/0000-0003-4487-823X

Author Contributions

Canan Yuksel-Haji: Conceptualization; Data curation; Methodology; Project administration; Writing – original draft; Writing – review & editing.
Nicola Hancock: Supervision; Writing – review & editing.
Allan Clark: Supervision; Writing – review & editing.
Valerie M. Pomeroy: Supervision; Validation; Writing – review & editing.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: VMP is supported by the NIHR (National Institute for Health and Care Research) HealthTech Research Centre in Brain Injury. The views expressed are those of the authors and not necessarily those of the NIHR or the Department of Health and Social Care.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

References

1. Kwakkel G, Lannin NA, Borschmann K, et al. Standardized measurement of sensorimotor recovery in stroke trials: consensus-based core recommendations from the Stroke Recovery and Rehabilitation Roundtable. *Int J Stroke*. 2017;12(5):451-461. doi:10.1177/1747493017711813
2. Burridge J, Alt Murphy M, Buurke J, et al. A systematic review of international clinical guidelines for rehabilitation of people with neurological conditions: what recommendations are made for upper limb assessment? *Front Neurol*. 2019;10:567. doi:10.3389/fneur.2019.00567
3. Pohl J, Held JPO, Verheyden G, et al. Consensus-based core set of outcome measures for clinical motor rehabilitation after stroke: a Delphi study. *Front Neurol*. 2020;11:875. doi:10.3389/fneur.2020.00875
4. Pollock A, St George B, Fenton M, Firkins L. Top 10 research priorities relating to life after stroke: consensus from stroke survivors, caregivers, and health professionals. *Int J Stroke*. 2012;9:313-320. doi:10.1016/S1474-4422(12)70029-7
5. Tang Q, Yang X, Sun M, et al. Research trends and hotspots of post-stroke upper limb dysfunction : a bibliometric and visualization analysis. *Front Neurol*. 2024;15:1449729. doi:10.3389/fneur.2024.1449729
6. Simpson LA, Hayward KS, McPeake M, Field TS, Eng JJ. Challenges of estimating accurate prevalence of arm weakness early after stroke. *Neurorehabil Neural Repair*. 2021;35(10):871-879. doi:10.1177/15459683211028240
7. Dalton EJ, Jamwal R, Augoustakis L, et al. Prevalence of arm weakness, pre-stroke outcomes and other post-stroke impairments using routinely collected clinical data on an acute stroke unit. *Neurorehabil Neural Repair*. 2024;38(2):148-160. doi:10.1177/15459683241229676
8. Mohammed S, Haidar J, Ayele BA, Yifru YM. Post-stroke limitations in daily activities: experience from a tertiary care hospital in Ethiopia. *BMC Neurol*. 2023;23(1):1-10. doi:10.1186/s12883-023-03419-9
9. Schnitzler A, Woimant F, Tuppin P, De Peretti C. Prevalence of self-reported stroke and disability in the French adult population: a transversal study. *PLoS ONE*. 2014;9(12):e115375. doi:10.1371/journal.pone.0115375
10. Knorr S, Rice CL, Garland SJ. Perspective on neuromuscular factors in poststroke fatigue. *Disabil Rehabil*. 2012;34(26):2291-2299. doi:10.3109/09638288.2012.683233
11. Levin MF. Understanding motor recovery and compensation in neurorehabilitation. In: Selzer M, Clarke S, Cohen L, Kwakkel G, Miller R, eds. *Textbook of neural repair and rehabilitation*. 2nd ed. Cambridge University Press; 2014:599-608. doi:10.1017/cbo9780511995583.044
12. Kwakkel G, Stinear C, Essers B, et al. Motor rehabilitation after stroke: European Stroke Organisation (ESO) consensus-based definition and guiding framework. *Eur Stroke J*. 2023;8(4):880-894. doi:10.1177/23969873231191304
13. Santisteban L, Térémetz M, Bleton JP, Baron JC, Maier MA, Lindberg PG. Upper limb outcome measures used in stroke rehabilitation studies: a systematic literature review. *PLoS ONE*. 2016;11(5):e0154792. doi:10.1371/journal.pone.0154792
14. Velstra IM, Ballert CS, Cieza A. A systematic literature review of outcome measures for upper extremity function using the international classification of functioning, disability, and health as reference. *PM R*. 2011;3(9):846-860. doi:10.1016/j.pmrj.2011.03.014
15. World Health Organization. *Towards a Common Language for Functioning, Disability and Health: ICF*. World Health Organization; 2002:1-22. Accessed April 14, 2023. http://www.who.int/classifications/icf/training/icfbeginnersguide.pdf
16. Xiong T, Hartley S. Challenges in linking health-status outcome measures and clinical assessment tools to the ICF. *Adv Physiother*. 2008;10(3):152-156. doi:10.1080/14038190802331401

17. World Health Organization. *How to Use the ICF: A Practical Manual for Using the International Classification of Functioning, Disability and Health (ICF). Exposure Draft for Comment*. WHO. Accessed April 14, 2023. https://cdn.who.int/media/docs/default-source/classification/icf/drafticfpracticalmanual2.pdf?sfvrsn=8a214b01_4&download=true
18. Merriam-Webster. Neuromuscular. 2011. Accessed April 16, 2023. https://www.merriam-webster.com/dictionary/neuromuscular
19. Salter K, Jutai JW, Teasell R, Foley NC, Bitensky J. Issues for selection of outcome measures in stroke rehabilitation: ICF body functions. *Disabil Rehabil*. 2005;27(4):191-207. doi:10.1080/09638280400008537
20. Salter K, Jutai JW, Teasell R, Foley NC, Bitensky J, Bayley M. Issues for selection of outcome measures in stroke rehabilitation: ICF activity. *Disabil Rehabil*. 2005;27(6):315-340. doi:10.1080/09638280400008545
21. Cieza A, Brockow T, Ewert T, et al. Linking health-status measurements to the International Classification of Functioning, Disability and Health. *J Rehabil Med*. 2002;34:205-210. doi:10.1080/165019702760279189
22. Cieza A, Geyh S, Chatterji S, Kostanjsek N, Üstün B, Stucki G. ICF linking rules: an update based on lessons learned. *J Rehabil Med*. 2005;37:212-218. doi:10.1080/16501970510040263
23. Demers M, Levin MF. Do activity level outcome measures commonly used in neurological practice assess upper-limb movement quality? *Neurorehabil Neural Repair*. 2017;31(7):623-637. doi:10.1177/1545968317714576
24. Gasperini G, Rota M, Guanziroli E, et al. Development and Rasch validation of an observational assessment tool of upper limb functional impairment in stroke survivors: Functional Assessment Test for Upper Limb (FAST-UL). *Arch Phys Med Rehabil*. 2023;104:597-604. doi:10.1016/j.apmr.2022.10.003
25. Wang S, Hsu CJ, Trent L, et al. Evaluation of performance-based outcome measures for the upper limb: a comprehensive narrative review. *PM R*. 2018;10(9):951-962. doi:10.1016/j.pmrj.2018.02.008
26. Van De Winckel A, Feys H, Van Der Knaap S, et al. Can quality of movement be measured? Rasch analysis and inter-rater reliability of the Motor Evaluation Scale for Upper Extremity in Stroke Patients (MESUPES). *Clin Rehabil*. 2006;20(10):871-884. doi:10.1177/0269215506072181
27. Turtle B, Porter-Armstrong A, Stinson M. A systematic review of the application and psychometric properties of the graded Wolf Motor Function Test. *Br J Occup Ther*. 2020;83(5):285-296. doi:10.1177/0308022619879074
28. Sivan M, O'Connor RJ, Makower S, Levesley M, Bhakta B. Systematic review of outcome measures used in the evaluation of robot-assisted upper limb exercise in stroke. *J Rehabil Med*. 2011;43(3):181-189. doi:10.2340/16501977-0674
29. Kim H, Shin JH. Assessment of upper extremity function in people with stroke based on the framework of the ICF: a narrative review. *Brain NeuroRehabil*. 2022;15(2):1-13. https://doi.org/10.12786/bn.2022.15.e16
30. Fritz SL, Blanton S, Uswatte G, Taub E, Wolf SL. Minimal detectable change scores for the wolf motor function test. *Neurorehabil Neural Repair*. 2009;23(7):662-667. doi:10.1177/1545968309335975
31. Taub E, Morris DM, Crago J. *Wolf Motor Function Test (WMFT) manual*. 2011. Accessed May 2, 2023. https://www.dpo.uab.edu/citherapy/images/pdf_files/CIT_Training_WMFT_Manual.pdf
32. Morris DM, Uswatte G, Crago JE, et al. The reliability of the wolf motor function test for assessing upper extremity function after stroke. *Arch Phys Med Rehabil*. 2001;82(6):750-755. doi:10.1053/apmr.2001.23183
33. Levin MF, Kleim JA, Wolf SL. What do motor "recovery" and "compensation" mean in patients following stroke? *Neurorehabil Neural Repair*. 2009;23(4):313-319.
34. Baker K, Cano SJ, Playford ED. Outcome measurement in stroke: a scale selection strategy. *Stroke*. 2011;42(6):1787-1794. doi:10.1161/STROKEAHA.110.608505
35. Sullivan JE, Crowner BE, Kluding PM, et al. Outcome measures for individuals with stroke: process and recommendations from the American physical therapy association neurology section task force. *Phys Ther*. 2013;93(10):1383-1396. doi:10.2522/ptj.20120492
36. Bernhardt J, Borschmann K, Boyd L, et al. Moving rehabilitation research forward: developing consensus statements for rehabilitation and recovery research. *Int J Stroke*. 2016;11(4):454-458. doi:10.1177/1747493016643851
37. Fugl-Meyer AR, Jääskö L, Leyman I, Olsson S, Steglind S. The Post-Stroke Hemiplegic Patient. I. A method for evaluation of physical performance. *Scand J Rehabil Med*. 1975;7:13-31.
38. Hervé-Colas J, Newton SP, Engelter ST, et al. Standardized international manual of the Fugl-Meyer Assessment of motor function after stroke. *Neurorehabil Neural Repair*. 2026. doi:10.1177/15459683251412300
39. Nakamura R, Moriyama S. *Rehabilitation Manual 8: Manual Function Test (MFT) and Functional Occupational Therapy for Stroke Patients*. 2020. Accessed December 20, 2025. https://www.rehab.go.jp/english/pdf/E08.pdf
40. Yozbatiran N, Der-Yeghiaian L, Cramer SC. A standardized approach to performing the action research arm test. *Neurorehabil Neural Repair*. 2008;22(1):78-90. doi:10.1177/1545968307305353
41. Homberg A. Conducting Delphi surveys in medical education research. *GMS J Med Educ*. 2026;43(1):1-23.